Measures and perceptual correlates of signal envelope changes induced by hearing aid compression

René Burmand Johannesson[1] and Justyna Walaszek[2]

[1] *Eriksholm Research Centre, Oticon A/S, Kongevejen 243, 3070 Snekkersten, Denmark*

[2] *Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark*

The motivation for this investigation was that envelope cues are important for understanding speech, whereas the compression system in a hearing aid will change the speech signal's envelope. Especially when the compression system is acting on a mix of speech and fluctuating noise, the changes of the speech signal envelope become complex. Three measures for quantifying different aspects of the induced envelope changes have been investigated: Across Source Modulation Correlation (ASMC), Envelope Difference Index (EDI) and the change of the long-term Signal to Noise Ratio at the hearing aid output (ΔSNR). These three measures and combinations of these were compared against the outcome of speech intelligibility tests performed on 14 normal-hearing listeners using vocoded speech and noises. The noises were all fluctuating with speech-like temporal properties. The speech material was based on the Dantale II test corpus and the masking noises on either ICRA two-talker modulated noise or a Danish audio book. The findings indicate how predictive the different measures are for speech intelligibility, and suggest that the effects of hearing aid compression on speech intelligibility are complex and cannot be covered by a single aspect of envelope changes.

THE ENVELOPE MEASURES

The three measures of envelope change were calculated using only input and output signals. This has the advantage that the measures can be obtained with any compression system in a non-invasive manner, because there is no need for knowledge about the signal levels within the compression system.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

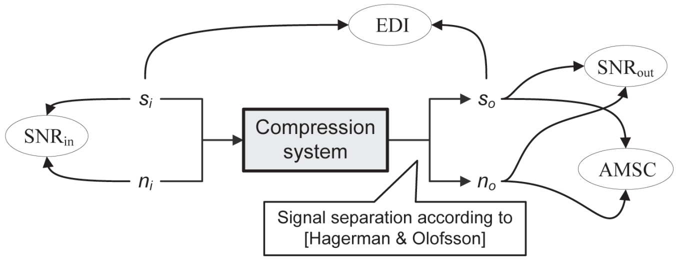

Fig. 1: Schematic of the signals used to derive the measures. The signals; s_i, n_i, s_o and n_o are speech and noise at the input and the output. The output signals s_o and n_o are separated using the method by Hagerman and Olofsson (2004). The long term SNRs at input and output, SNR_{in} and SNR_{out}, are used to determine ΔSNR. ASMC is based on s_o and n_o, EDI is based on s_i and s_o.

Across Source Modulation Correlation (ASMC)

This measure quantifies interaction between compression system and the speech and noise envelopes and was introduced by Stone and Moore (2007). A higher absolute value of ASMC expresses an increased common modulation of speech and noise through the compression system, which could cause perceptual fusion of speech and noise.

$$\text{ASMC} = \frac{1}{K}\sum_{k=1}^{K} r\left(a_k, b_k\right) \qquad \text{(Eq. 1)}$$

where a_k, b_k is logarithm of the envelopes of output speech and noise signals (s_o and n_o) in analysis channel number k of K frequency analysis channels and r is Pearson's correlation.

Envelope Difference Index (EDI)

This measure quantifies the difference between the temporal speech envelopes at the input and at the output of the compression system. The measure was introduced by Fortune *et al*. (1994) and was in this investigation modified to cover the envelope in multiple frequency channels. The EDI number expresses an amount of change in envelope shape, which could deteriorate speech perception.

$$\text{EDI} = \frac{1}{2KM}\sum_{k=1}^{K}\left(\sum_{m=1}^{M}\left|env1_{km} - env2_{km}\right|\right) \qquad \text{(Eq. 2)}$$

where $env1$ is the envelope of the unprocessed signal (s_i), $env2$ is the envelope of the processed signal (s_o) and m is the sample number of $1 \ldots M$ samples analysed in k of K frequency channels.

Change in long term SNR (ΔSNR)

This measure, employing the method and system described in Naylor and Burmand Johannesson (2009), quantifies the effect of the compression system on long-term speech and noise levels. The ΔSNR measure indicates whether the compression system increases or decreases the long-term SNR, which could affect speech perception.

$$\Delta SNR = SNR_{out} - SNR_{in} \qquad \text{(Eq. 3)}$$

where SNR_{in} is the long-term SNR at compression system input and SNR_{out} is the long-term SNR at compression system output.

An investigation by Naylor *et al.* (2008) found significant performance changes which for hearing impaired subjects mainly could be explained by ΔSNR. However for normal hearing subjects the perceptual performance change could not be attributed to ΔSNR.

EXPERIMENT

Design background

The focus was on envelope cues and how they are affected by a compression system. The objective was to maximize the variance in effects expected to contribute to ASMC and EDI, while preventing ΔSNR from becoming a dominant factor. These basic design rules lead to the following choices:

Vocoded signals

The presented speech and noise material was processed using a 'noise vocoder' removing the temporal fine structure from the signals (Hopkins and Moore, 2007), thus the test subjects were provided mainly with envelope cues.

Noise with speech-like modulations, and SNR close to zero

At $SNR_{\iota\nu}$ levels from 0 dB to ±5 dB, only small ΔSNR effects will be observed when speech and noise envelopes contain similar amounts of modulation (Naylor and Burmand Johannesson 2009). At the same time, these conditions ensure relatively large values of ASMC and EDI. Furthermore the modulation in speech and noise contribute to the absolute values of the ASMC and EDI measures.

Fast and slow compression applied commonly and independently

Realistic compression system differences were obtained by contrasting fast and slow release times. An accentuated contrast was obtained by applying compression either to the mixture of speech and noise (Common) or to the speech and noise signals independently before mixing (Indep). The Indep condition cannot be realised in practical compression systems, but gives experimental contrast as ASMC is always asymptotically zero since the sources do not interact.

Experiment: Implementation

Test subjects: 14 (3 females and 11 males) normal-hearing (<15 dB HL) aged from 21 to 30 years. All test subjects received a gift as compensation for their participation.

Speech material: Dantale II material (5-word Hagerman-type sentences Wagener *et al*., 2003).

Noise material: ICRA (Dreschler *et al*., 2001) 2-talker modulated noise (N2spk) or speech from a Danish audio book (N1spk).

Parameters: The test varies 3 parameters; the compression speed, the mixing of signals before or after compression and finally the type of noise.

This gives: [Fast/Slow]•[Common/Indep]•[N2spk/N1spk] ⇒ 8 test cases.

Training: The test task is quite difficult and in order to reduce confounding training effects the test subjects had 2 visits: training and then the test. The training was separated into 5 steps of increasing difficulty totalling 220 sentences.

1. Introduction of the speech material with 10 sentences of unprocessed speech signal.
2. Presentation of the effect of noise vocoder processing on 10 sentences.
3. The easy task with a mixture of vocoded speech and noise at a very high SNR (approximately 20 dB), 10 sentences for each of the 2 noise types.
4. Introduction to the actual test at an intermediate difficulty level with a high SNR (approximately 6 dB), 15 sentences was used for each of the 8 test cases.
5. Presentation of the test subject to the real test. 5 sentences for each of the 8 test cases.

Test: The test was made at a SNR_{in} of +4.0 dB for N2spk and +4.5 dB for N1spk. These SNR_{in} values were determined as producing 70% correct on average for uncompressed vocoded speech and noise. For all of the 14 subjects, each of the 8 test cases were tested with 30 sentences (totalling 240 sentences). The sentences were presented diotically via ER2 insert earphones at 65 dB SPL.

Compression system: 15 frequency channels with constant CR=2.5 over the full input dynamic range and an attack time of 15 ms and release times of 80 ms or 1,000 ms.

RESULTS

The results of the perceptual experiment are shown in Table 1; the reported % correct score is the average over all test subjects.

	Slow Indep N_{2spk}	Slow Common N_{2spk}	Fast Indep N_{2spk}	Fast Common N_{2spk}	Slow Indep N_{1spk}	Slow Common N_{1spk}	Fast Indep N_{1spk}	Fast Common N_{1spk}
ASMC	0.020	-0.049	0.009	-0.230	0.003	-0.034	-0.005	-0.140
EDI	0.13	0.11	0.18	0.16	0.16	0.14	0.26	0.22
ΔSNR	-3.28	-1.58	-1.73	-2.18	-0.83	0.28	-0.82	0.06
% correct	62.5	65.6	64.6	57.2	66.8	70.9	61.0	61.1
95% confidence	*±6,6*	*±7,0*	*±6,5*	*±6,5*	*±7,3*	*±7,8*	*±9,2*	*±7,3*

Table 1: Summary of experimental results.

The ASMC data range [-0.23; 0.02] covers what can be found in Stone and Moore (2007). The EDI range [0.11; 0.26] is a bit less than reported by Fortune *et al*. (1994) and finally ΔSNR [-3.3; 0.3] is about half of the range of values investigated by Naylor and Burmand Johannesson (2009). Thus the experimental design objective stated previously was achieved.

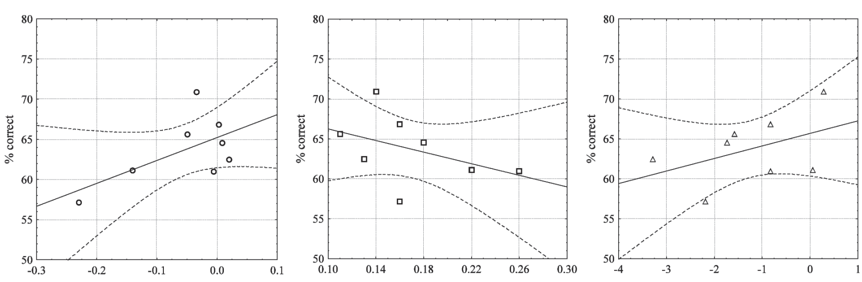

Fig. 2: The average of % correct plotted against the three measures ASMC (○), EDI (□) and ΔSNR (Δ), the solid lines show the best linear fits and the dotted lines show the 95% confidence intervals.

The three individual envelope measures are only moderately correlated to the observed % correct. (ASMC: R=0.593, EDI R= -0.423 and ΔSNR R= 0.438).

Table 2 shows the correlations between the individual envelope measures and subsets of the observed % correct for each of the experimental contrasts.

	Slow/ Fast	N_{2spk}/ N_{1spk}	Indep / Common
ASMC	0.466	-0.113	**0.731***
EDI	**-0.759***	-0.542	0.271
ΔSNR	-0.084	**-0.845***	-0.367

Table 2: Correlations between the three measures ASMC, EDI and ΔSNR and the % correct subdivided according to the parameters varied in the experiment. Significant ($p<0.05$) correlations are marked with *.

Only three significant ($p<0.05$) correlations were found, one for each of the three measures:

- ASMC correlates with the Indep/Common contrast. This complies with the assumption that independently or commonly applied compression would give experimental contrast for envelope interaction effects.
- EDI correlates with effect of the speed of the compression system: Slow/Fast. This finding shows that the change of the envelope shape is mostly determined by the speed of the compression system.
- ΔSNR correlates with effect of the noise type: N2spk/N1spk. This is in accordance with the findings in Naylor and Burmand Johannesson (2009). We might expect that ΔSNR also would be correlated with effect of the speed of the compression system, but the selected SNR_{in} range renders this undetectably small.

The following correlation values in Table 3 show that the three measures are reasonably uncorrelated and consequently may offer a good basis for a regression analysis.

Measure A	Measure B	Correlation between A & B
ASMC	ΔSNR	-0.063
EDI	ΔSNR	0.392
ΔSNR	EDI	-0.076

Table 3: Correlation between the three measures ASMC, EDI and ΔSNR (all are non-significant)

Regression analysis

The regression analysis modelled the speech intelligibility in % correct as a function of ASMC, EDI and ΔSNR combined. The resulting model can be expressed as:

$$\% \text{ Correct}_{\text{estimated}} = b0 + b1 \cdot ASMC + b2 \cdot EDI + b3 \cdot \Delta SNR \qquad \text{(Eq. 4)}$$

Where b0= 78.54, b1=29.20, b2=-58.21 and b3=2.689

This regression model is highly correlated with test subjects' scores:

R= 0.998, Adjusted R^2= 0.96 ($p< 0.002$).

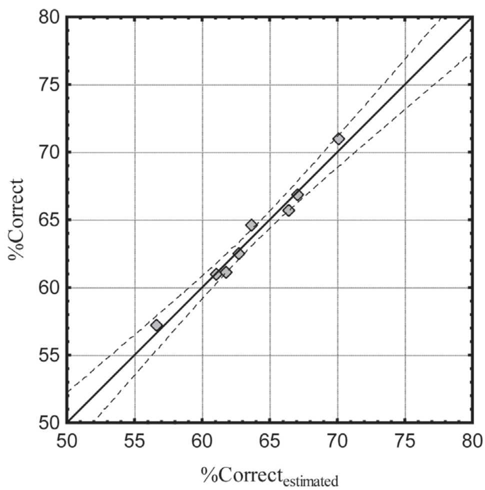

Fig. 3: Test subject average from table 1 plotted against the regression model, the observed correlation is 0.998. The solid lines show the linear regression fit and the dotted lines the 95% confidence interval.

Removing any of the three measures ASMC, EDI or ΔSNR from the regression model gives much poorer prediction of the observed scores, and non-significant results in all cases:

Leaving out EDI and using ASMC and ΔSNR: R= 0.766, adjusted R^2= 0.42 ($p< 0.11$).

Leaving out ΔSNR and using ASMC and EDI: R= 0.707, adjusted R^2= 0.30 ($p<0.18$).

Leaving out ASMC and using ΔSNR and EDI: R= 0.781, adjusted R^2= 0.46 ($p<0.09$).

CONCLUSIONS

The three objective measures ASMC, EDI and ΔSNR quantify different side-effects that compression may have on the speech envelope. These side-effects are (i) interaction of speech and noise envelopes, (ii) changes in speech envelope shape and (iii) changes of overall speech and noise levels. This experiment shows that a perceptual measure like speech intelligibility may be sensitive to each of these side-effects. The objective measures are weakly correlated with each other, but a combination of the objective measures may provide a good estimate of the perceptual measure. This suggests that studies seeking to explain the effects of compression on speech intelligibility should be designed with careful attention to which side-effects of compression in fact will be operative. Since the present experiment was deliberately designed to provide maximum independent variation of the envelope measures ASMC, EDI and ΔSNR, it would be premature to conclude that these represent a necessary and sufficient set of measures to describe compression effects. Furthermore, the data set is small, so the multiple regression model derived above should be regarded as demonstrative rather than predictive in a general way.

ACKNOWLEDGEMENTS

We would like to give our thanks to Torsten Dau, Filip Munch Rønne and Niels Henrik Pontoppidan for valuable input to this investigation.

REFERENCES

Stone, M. A., and Moore, B. C. J. (**2007**). “Quantifying the effects of fast-acting compression on the envelope of speech,” J. Acoust. Soc. Am. **121**, 1654-1664.

Fortune, T. W., Woodruff, B. D., and Preves, D. A. (**1994**). “A new technique for quantifying temporal envelope contrasts,” Ear Hear **15**, 93-99.

Naylor, G., and Burmand Johannesson, R. (**2009**). “Long-Term Signal-to-Noise Ratio (SNR) at the Input and Output of Amplitude Compression Systems,” J Am Acad Audiol **20**, 161-171.

Naylor, G., Rønne, F. M., and Burmand Johannesson, R. (**2008**). “Perceptual correlates of the long-term SNR change caused by fast-acting compression” ICHON 2008 poster presentation.

Hagerman, B., and Olofsson, Å. (**2004**). “A Method to Measure the Effect of Noise Reduction Algorithms Using Simultaneous Speech and Noise,” Acta Acustica **90**, 356–361.

Hopkins, K., and Moore, B. C. J. (**2007**). “Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information,” J. Acoust. Soc. Am. **122**, 1055–1068.

Wagener, K., Lignel Josvassen, J. and Ardenkjær, R. (**2003**). “Design, optimization and evaluation of a Danish sentence test in noise,” International Journal of Audiology **42**, 10-17.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). “ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment,” Audiology **40**, 148-157.